



TERA Consultants

39, rue d'Aboukir

75002 PARIS

Tél. + 33 (0) 1 55 04 87 10

Fax. +33 (0) 1 53 40 85 15

www.teraconsultants.fr

S.A.S. au capital de 200 000 €

RCS Paris B 394 948 731

Big Data : concurrence et consommation

Laurent Benzoni

Laurent.benzoni@u-paris2.fr

benzoni@tera.fr

19 octobre 2016

Un développement d'Internet...



Phase **1** : Internet **universitaire** (Arpanet, initiale)

Phase **2** : Internet **vitrine** (présence statique sur Internet)

Phase **3** : Internet **transactionnel** (e-commerce, www + .com, etc.)

2003, Nicholas Carr annonce dans un article très remarqué de la Harvard Business Review :
« Les technologies d'information sont banalisées, elles ne sont plus un enjeu stratégique ».

Phase **4** : Internet **social** et **collaboratif** (en cours actuellement)

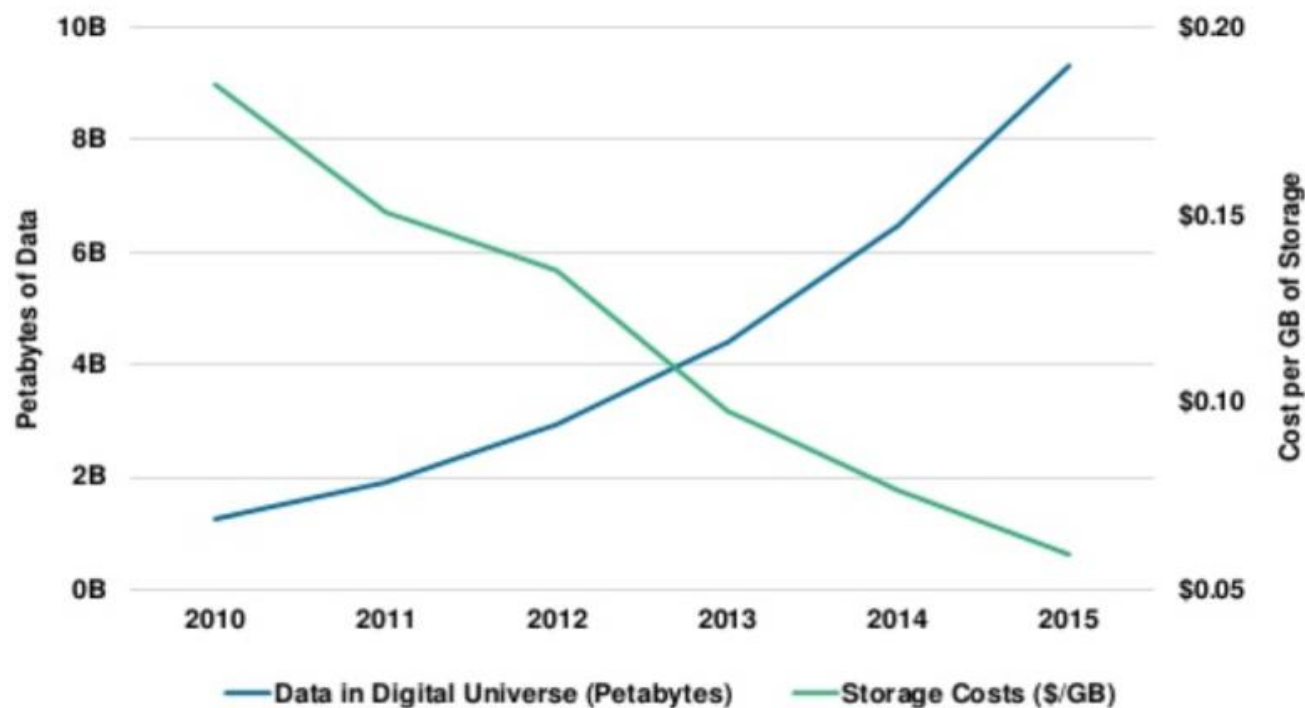
Phase **5** : Internet **sensoriel** ou **Internet des Objets** (capteurs température, pression, vibration, luminosité, humidité, tension, etc.)

Phase **6**

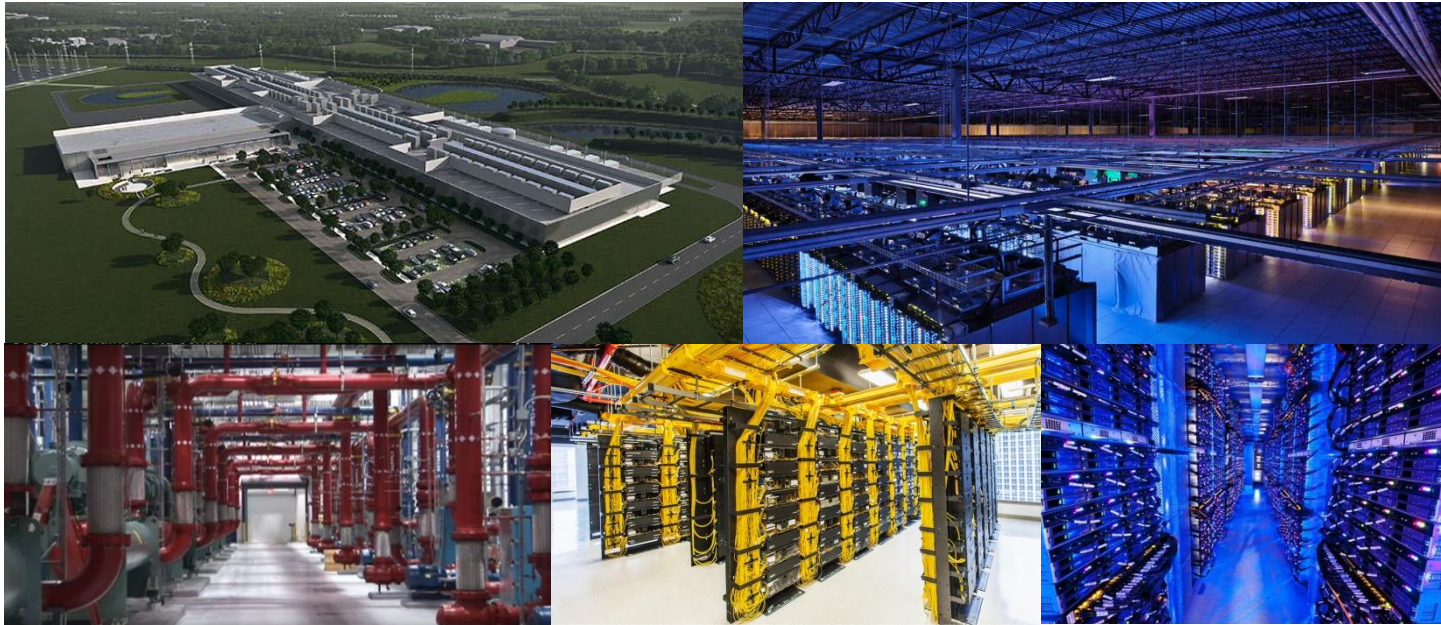
... en lien avec la croissance des données stockées



Evolution des volumes de données stockées et du coût de stockage



Derrière le cloud, des infrastructures...



Des services avec un sous-jacent industriel et capitalistique

Investissements dans les infrastructures en 2014

Google : 11 milliards de dollars

Amazon : 6 milliards de dollars

Microsoft : 5,3 milliards de dollars

Facebook... seulement 1,8 milliard de dollars

Position sur le marché des prestations du cloud

- 1. Amazon Web Services (31%)**
- 2. Microsoft (9%)**
- 3. IBM (7%)**
- 4. Google (4%)**

Services en réseau et phénomène d'externalité

Propriété 1/ L'utilité d'un service en réseau pour un utilisateur est d'autant plus grande que le nombre d'utilisateurs de ce service est grand.

Propriété 2/ La décision d'un utilisateur d'utiliser un service en réseau dépend du nombre d'utilisateurs du service et sa décision d'adhérer accroît l'utilité des utilisateurs existants.

Interdépendance des décisions → EXTERNALITES

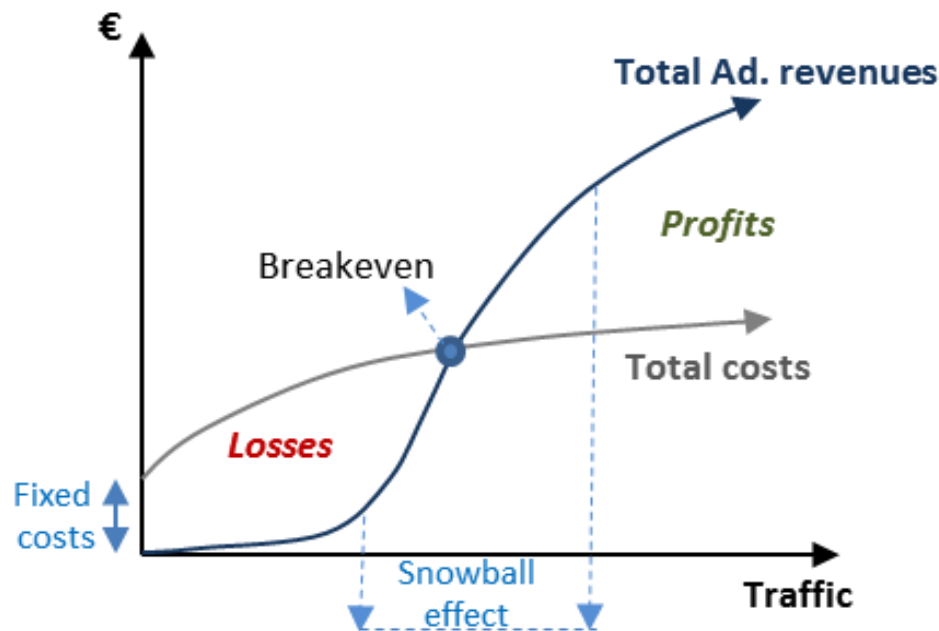
Exemples de services



Mais qui se connecte à un nouveau service en réseau, donc sans utilisateurs ?

Produire la taille critique initiale

Gratuité et facilité d'usage sont logiquement des facteurs clés de l'amorçage de la base initiale des utilisateurs → Le modèle standard du « tout gratuit »



Une fois l'amorçage enclenché, la croissance s'autoalimente (effet boule de neige, crowd effect). Un service s'impose et domine son marché : **«winner take all»**..... tout en restant menacé par de nouveaux services substitués offrant d'autres fonctionnalités.

Externalités croisées et marchés biface : valoriser les données acquises gracieusement sur d'autres marchés

Du concept d'externalités positives croisées...

LJ. Meade (prix Nobel d'économie 1977) crée le concept d'**externalité positive croisée** (1952) et l'illustre par la relation entre apiculteurs et arboriculteurs.

- *L'apiculteur* bénéficie de la proximité de l'arboriculteur pour produire son miel en quantité et en qualité. *L'arboriculteur* génère une externalité positive pour l'apiculteur.
- *Symétriquement*, l'arboriculteur profite de la pollinisation de ses arbres et bénéficie d'une externalité positive générée par l'apiculteur.
- *L'externalité se réalise en deux sens : elle est croisée. Il faut optimiser-maximiser cette relation croisée pour maximiser le surplus social.*

... au concept de marchés bi ou multi-face.

- Le concept de **marchés biface** apparaît au début des années 2000.
- Il se réfère aux activités, dites de «plateformes d'intermédiation», valorisant des externalités positives croisées en adressant simultanément deux marchés ou plus constituant les faces de la plateforme. Exemples : médias (intermédiation entre lecteurs et annonceurs), cartes de crédit (intermédiation entre consommateurs 'encartés' et commerçants), foires, marchés boursiers, etc.

La tarification de chaque face intègre leur interdépendance...

Principe de la gratuité

Condition nécessaire : la plateforme du service en réseau dispose d'au moins deux groupes distincts d'utilisateurs ou "faces" qui produisent des externalités positives mutuellement positives à chacun d'entre eux.

Principe de tarification

Les utilisateurs de la face ayant la plus forte sensibilité aux prix (internauts) est tarifiée à un prix inférieur au coût marginal, le prix peut même être nul (gratuité du service) ou négatif (subvention). Les utilisateurs sur l'autre face du marché ayant une élasticité au prix plus faible (annonceurs) sont tarifés à un prix supérieur au coût marginal → existence de subventions croisées entre les deux faces (utilisateurs)

Exemples : tout modèle économique basé sur le financement par la publicité (services online, médias), cartes bancaires, jeux vidéo/ consoles mais pas le modèle imprimantes/ cartouches d'encre

Conditions de mise en oeuvre

- Niveau d'externalité générée par chaque face
- Niveau de l'élasticité de la demande aux prix sur chacune des faces
- Coût de mise en oeuvre de la subvention et surveillance



Mais, si c'est gratuit, c'est vous le produit...

Données individuelles et valorisation par les services en réseau



Les plateformes de service disposent de **trois types d'informations** sur leurs utilisateurs

- 1 - L'information déclarée** : transférée explicitement par les utilisateurs à travers des formulaires, questionnaires, etc.
- 2 - L'information observée** : transférée tacitement par l'usage du service : pages et sites visités (cookies), requêtes, achats, etc.
- 3 - L'information inférée** : produite par la plateforme à partir des deux types informations précédentes.

La base des données et les **algorithmes** produisent l'information inférée qui enrichit la profondeur et la qualité de la connaissances des usagers donc la valeur des données. *Plus les informations déclarées et observée sont importantes, plus l'information inférée est importante.*

La plateforme s'enrichit constamment de nouveaux services plus ou moins liés au service initial afin que la base des utilisateurs accroissent leur présence et leurs échanges. Les algorithmes affinent les préférences des individus pour les inciter à consommer ce qui est proposé sur la plateforme : c'est l'effet « **bubble filter** »

Margrethe Vestager : « **Plus vous collectez des données, plus vous savez le meilleur produit que vous pouvez fournir, mais aussi plus votre pouvoir de prescription augmente, ce qui peut poser un problème de concurrence** » MLex Interview: Margrethe Vestager, MLex Special Report, 22

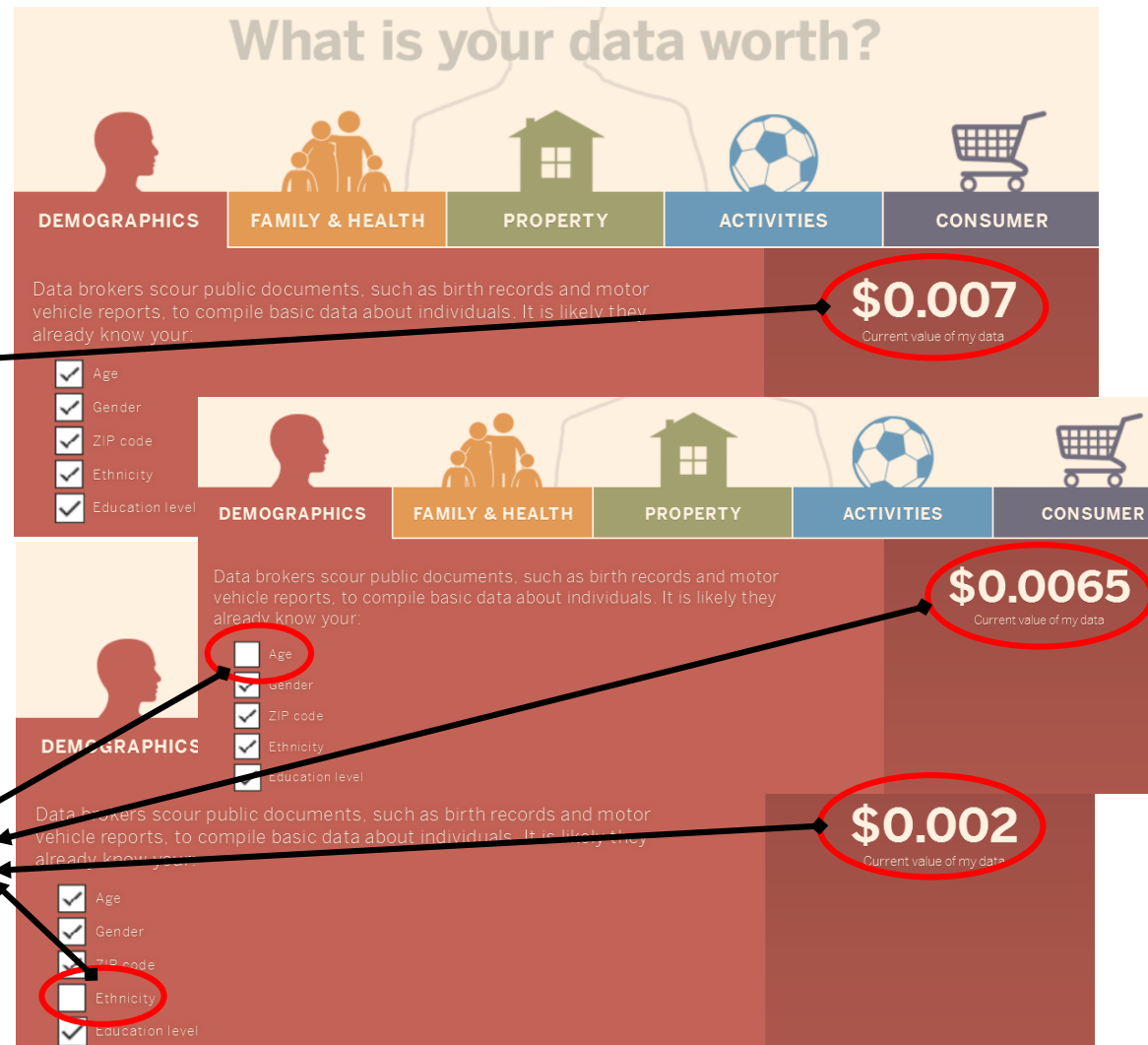
January 2015 ('Vestager Interview'), <http://mlexmarketinsight.com/wp-content/uploads/2015/01/MLex-Interview-Vestager-22-01-151.pdf>.

Quelques prix liés aux valeurs commerciale et/ou publicitaire de données individuelles

Plus de 270 brokers de données dans le monde. Il existe ainsi un prix de marché des données individuelles.

- Données générales sur un individu : **0,007ct** ;
- Millionnaire : **0,123ct** ;
- Allergique : **0,577ct** ;
- Asthme ou arthrose : **1,037\$....**

NB. Le critère d'appartenance ethnique d'un individu est valorisée dix fois plus que la connaissance de son âge...



Source :

http://www.ft.com/cms/s/2/927ca86e-d29b-11e2-88ed00144feab7de.html?ft_site=falcon#axzz2WfFmKwic

De la valeur individuelle à la valeur sociale.

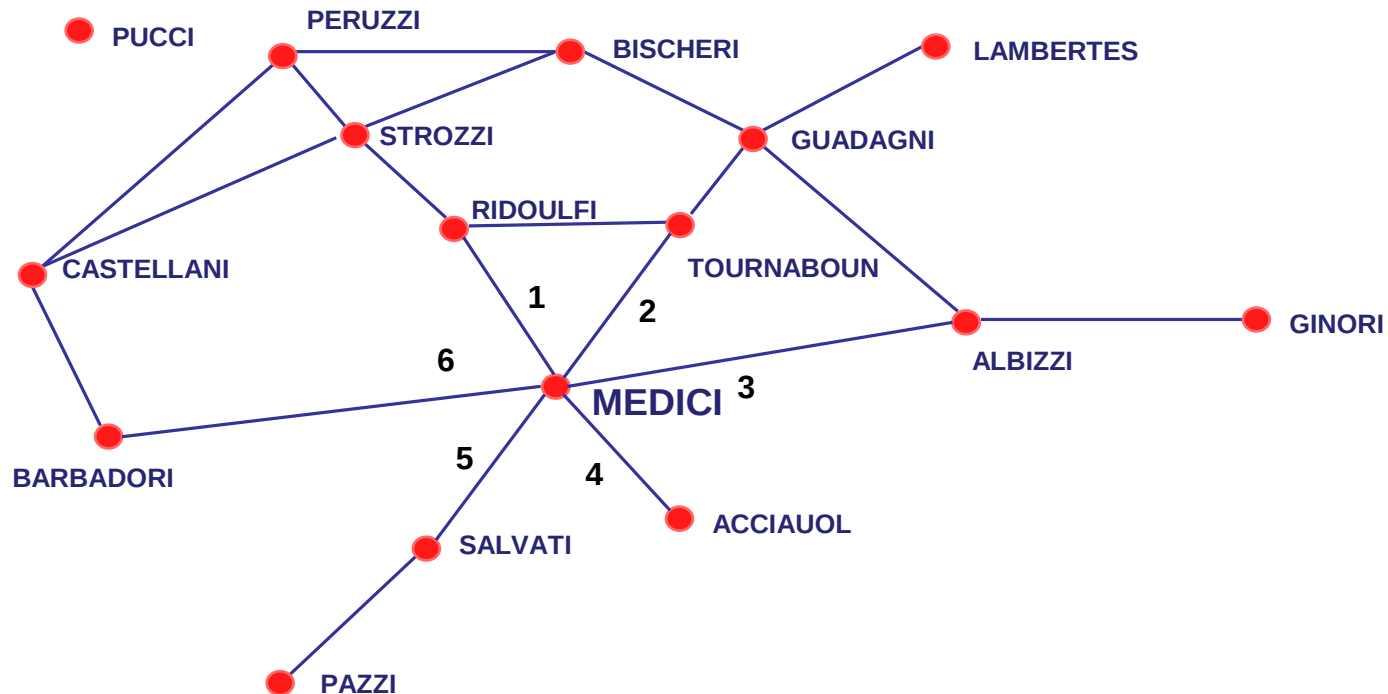


Source : <https://www.cashinfo.com/infos/calculateur-de-valeur>

L'individu dans son « graphe social »

XVème siècle : Florence est dirigée par une oligarchie de 16 familles les plus influentes au conseil de la ville. Les *Médici* ne sont pas les plus riches, ni les mieux représentés au conseil... Pourtant, ils deviendront l'une des familles les plus influentes de Florence. **Explication par le graphe social...**

L'influence des Medici réside dans les liens créés par le mariage. Une **famille** est représentée par un **point rouge** : un **nœud** en théorie des graphes. Un **lien** (en **bleu**) indique un **mariage** entre deux familles. Le nombre de liens partant d'un nœud mesure son **degré de densité**.

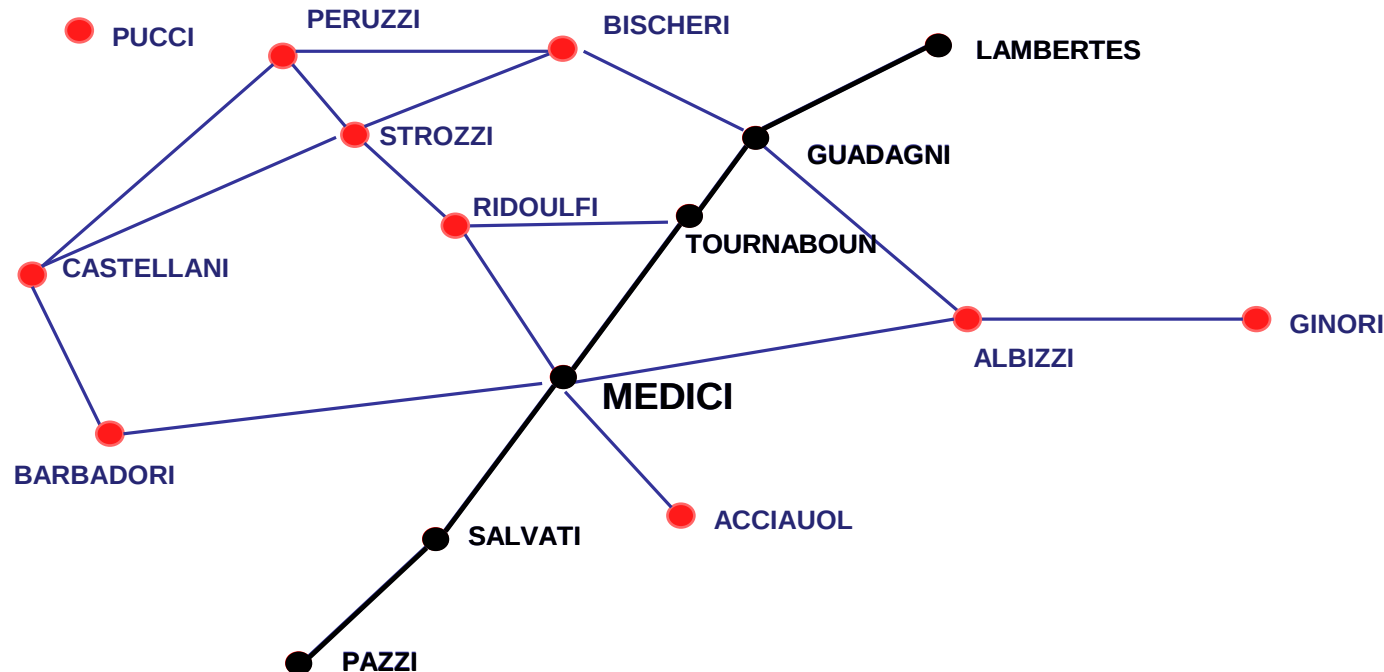


Les Medici : nœud de densité 6 contre 4 au maximum pour les autres familles. Dans les réseaux sociaux d'aujourd'hui, un nœud est un individu, un lien est un contact de 1^{er} rang (ex. LinkedIn). Le degré de densité d'un individu est son nombre de contacts de 1^{er} rang : **signe apparent de son influence**.

Le diamètre du réseau : la dimension du « monde »



Le diamètre d'un réseau est la distance (nbre de liens) séparant les deux nœuds les plus éloignés. Il mesure les degrés de séparation (ligne noire sur schéma ci-dessous). Plus le **nombre de degré de séparation est petit**, plus la distance moyenne à parcourir entre les nœuds est faible, plus le **réseau est potentiellement viral**. Les amis...de mes amis... de mes amis...



Milgram (1967) mesure le nombre moyen de degré de séparation entre deux individus tirés au hasard aux USA : il en trouve 6 (5,2). **La théorie du « petit monde » est née.**

Mesurer sur 721 millions d'utilisateurs de Facebook : le monde est plus petit, **4,74 degrés** de séparation – soit moins de 5 personnes – entre deux individus pris au hasard sur le réseau social. Il passe à 3 si l'analyse est circonscrite à un seul pays. (cf. <https://arxiv.org/abs/1111.4570>)

La centralité : une mesure d'influence



La **densité** d'un réseau mesure le nombre de liens effectifs par rapport aux liens possibles. Pour le réseau des 16 familles florentines dominantes, il y a $16 * 15 = 340$ liens possibles et 19 liens effectifs de mariage, la densité du réseau est donc de $19/340 = 0,0558$.

Elle est bien plus faible sur Facebook par exemple car les 721 millions de comptes Facebook recensés pour l'expérience de Milgram étaient reliés par seulement 69 milliards de liens..

La **centralité d'intermédiation** d'un nœud évalue sa situation et sa capacité de mise en relation avec les autres nœuds : c'est une **mesure d'influence**. Dans un réseau social, elle mesure en quelque sorte la fraction de l'information du réseau qui passe systématiquement par l'individu dont on évalue la centralité.

$$B_k = \sum_{i,j} \frac{P_k(i,j)/P(i,j)}{(n-1)(n-2)/2}$$

L'intermédiation des Medici était de 0,522. La centralité d'intermédiation de la deuxième famille florentine après les Medici n'affichait qu'une valeur de 0,255, soit deux fois moins que les Medici. On comprend ainsi comment et combien les liens de mariages tissés avec les autres familles ont placé les Medici en position d'influence. Ils ont ainsi pu asseoir leur pouvoir sur Florence.

La mesure du graphe social constitue un ingrédient essentiel de la connaissance et de la valorisation des données personnelles recueillies par les services en réseau.

NB. La connaissance et maîtrise du graphe social sont très couteuses en capacités de stockage et très exigeantes en capacité de traitement analytique des données recueillies.

Le big data pour mesurer les réseaux d'influence

« Avoir des amis, c'est avoir du pouvoir. »

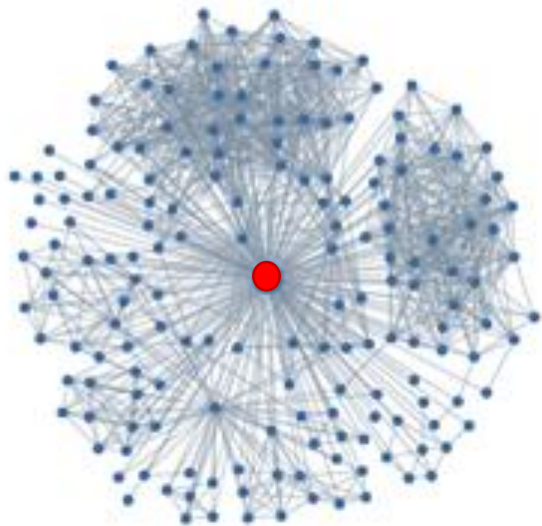
Thomas Hobbes, Léviathan , 1651

Traité de la matière, de la forme et du pouvoir d'une république ecclésiastique et civile

Quelle influence dans un graphe social

Un utilisateur Facebook affiche en moyenne 394 amis

Amis



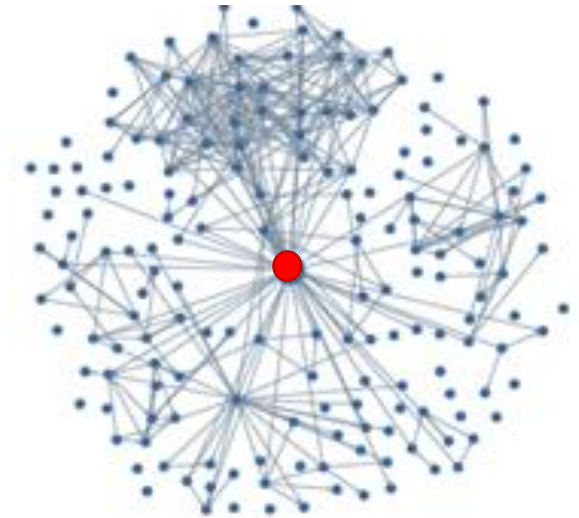
Liens de toute nature de degré 3 d'un utilisateur de réseau social



Impression de forte densité

La nature d'une relation peut se mesurer le nombre et le sens des messages échangés. En effet, les échanges sont coûteux en temps : on consacre du temps aux relations qui comptent. Pour mesurer l'intensité d'une relation entre deux noeuds, on peut filtrer les liens en mesurant la fréquence et le sens des échanges.

Amis ou contacts ?



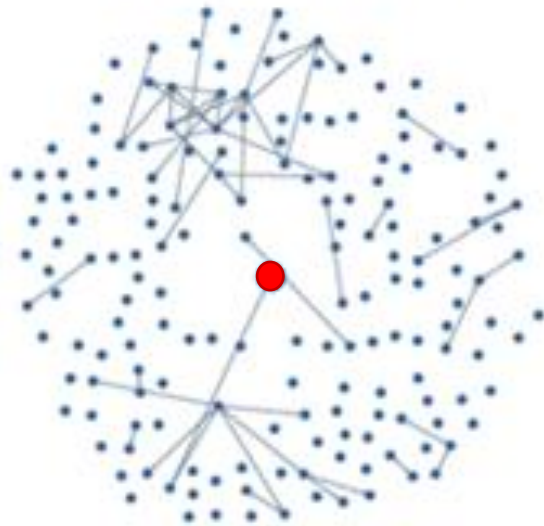
Liens avec **échange** de communication au moins une fois par trimestre



La densité se clairseme

« Vrais » amis, influenceurs

Les vrais amis ?



Liens avec échange de communication au moins une fois par jour



Une faible densité



Mais, un réseau social n'est-il qu'un moyen de communication ? Ou bien est-il aussi un moyen de diffusion

150 : nombre de Dunbar

Nombre maximal de personnes avec lequel un individu peut entretenir une relation suivie et stable. Cette limite serait inhérente à la capacité de notre cerveau à entretenir des liens affectifs.

Les influenceurs



Messages monodirectionnels sans rebond



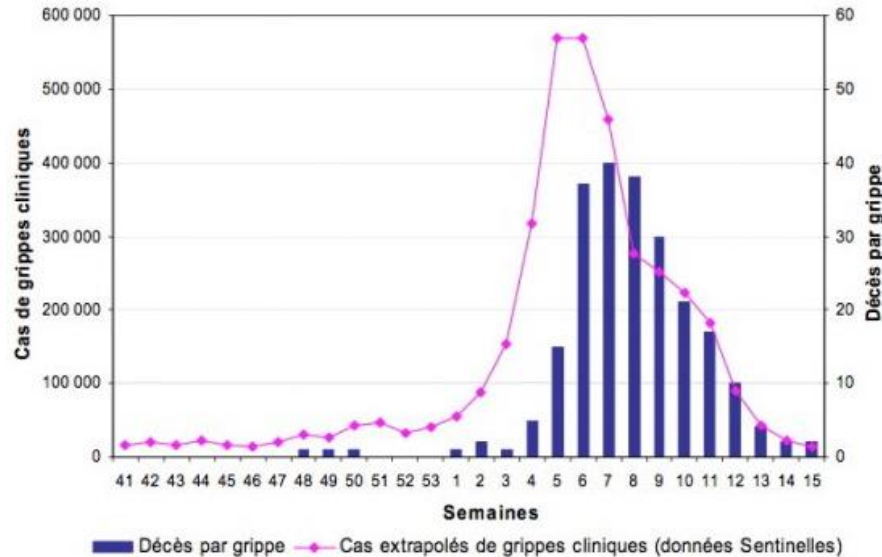
Cristiano Ronaldo

- 104,7 millions d' amis sur Facebook
- 47,3 millions de followers sur Twitter
- Le tweet sponsorisé vaut 230 000€

Tout est viral : dans quelle proportion ? Avec quelle prévisibilité ?

La modélisation type de la viralité par

La fonction de propagation (exemple d'une épidémie de grippe*)



Trois variables clé de la propagation :
vitesse, ampleur et durée.

NB. Une forte corrélation entre les requêtes sur les maladies sur Google et la propagation des maladies a été observée : cf.

<https://websenti.u707.jussieu.fr/sentiweb/?page=google>

Prévoir et provoquer la viralité sur les réseaux sociaux ?

Probabilité forte d'autant plus forte de viralité que l'information initiale est lancée ou relayée par un(des) influenceurs (Youtube : ice bucket challenge par exemple).

Mais, nombre de phénomènes viraux importants ne partent pas d'influenceurs identifiés (Youtube : Gagnam style, Harlem shake,...)

A défaut de prévoir, il faut anticiper par l'analyse en temps réel de propagation : seul celui qui observe les données au plus près le phénomène peut réaliser cette opération et... la valoriser.

Question de concurrence : valoriser les données acquises dans les opérations de concentration ?

Si les données sont à la base de la puissance de marché en ce sens qu'elle procure une valeur et un avantage concurrentiel, la face des plateformes où sont collectées gracieusement les données (internauts) constituent-elles un marché ?
Selon la jurisprudence, il n'existe pas un marché de l'audience mais un(des) marchés de la publicité par média.

La valeur des transactions est-elle un indice de la puissance de marché acquise* ?

Société	Date	Acheteur	Prix \$	Users	C.A. \$	Users à date
Instagram	Avril 2012	Facebook	1 Md	30 M	2M	500 M (06/2016)
YouTube	Nov 2006	Google	1,65 Md	50 M	15 M	Plus 1 Md (07/2016)
Minecraft	Sept. 2014	Microsoft	\$2,5 Md	100 M	14,3 M	40 M (06/2016)
WhatsApp	Octobre 2014	Facebook	\$21,6 M	430 M	15 M	Plus 1 Md (08/2016)
LinkedIn	Juin 2016	Microsoft	\$26,2 Md	433 Mn	3 Md	

* M. Vestager described data as 'The new currency of the Internet' during hearings before the European Parliament. Source : James Kanter, 'Antitrust Nominee in Europe Promises Scrutiny of Big Tech Companies', New York Times, 3 October 2014.

Exemple actuel de constitution d'une position de marché par croissance externe sur un service « big data »

Une personne moyenne génère plus de 1 million de giga-octets de données médicales au cours de sa vie. Avec Watson Health Cloud, IBM veut créer un écosystème pour la santé : aide au diagnostic, prévention, réduction des coûts, assurance maladie, etc. Les acquisitions de compétences et de données sont réalisées par croissance externe.



- Mars 2015
- Prix non divulgué
- Analyse en cloud des soins
- Avril 2015
- Prix non divulgué
- Santé de la population
- Octobre 2015
- \$1 milliard
- Plateforme d'imagerie médicale
- Octobre 2016
- \$2,6 milliards
- 215 millions de dossiers médicaux (1 américain sur 3)

Question de concurrence : croisement des données et renforcement de la puissance de marché ?

Le croisement des données doit-il être considéré comme un facteur d'accroissement de la puissance de marché ? Peut-il être sous le contrôle des autorités de la concurrence ?

ex. Facebook-Whatsapp

*"Nous allons explorer des façons dont les **entreprises et vous pouvez communiquer ensemble [...],** comme passer des commandes, effectuer des transactions, obtenir des informations de rendez-vous, recevoir des bons et avis de livraisons, être informé des mises à jour de produits et services et du marketing [...]"*

Nous n'autorisons toujours pas les bannières de publicité de tiers indépendants [...]
*Les messages envoyés par les marques pourront contenir une offre pour quelque chose qui pourrait vous intéresser. **En nous coordonnant davantage avec Facebook, nous pourrons faire des choses telles que suivre des mesures de base sur la fréquence d'utilisation de nos services des gens et améliorer la lutte contre les spams sur WhatsApp. Et en connectant votre numéro de téléphone avec les systèmes de Facebook,** ce dernier peut vous offrir de **meilleures suggestions d'amis et vous montrer des publicités plus pertinentes si vous avez un compte Facebook.***

Question de concurrence : certaines segmentations en marchés pertinents sont-elles encore pertinentes?



Les réseaux sociaux et les services de mise en relation entre des producteurs de contenus et des consommateurs de contenus sont « hybrides » dans le sens où ils sont autant moyens de communication que moyens de diffusion, c'est-à-dire des médias ?

Pourtant les médias sont soumis à des obligations spécifiques. Un traitement inégal pour des fonctionnalités identiques n'emporte-t-il pas des effets anticoncurrentiels ?



Question de concurrence : à qui appartiennent les données inférées ?

Les données inférées sont-elles la propriété de l'individu ou bien la propriété de l'entreprise qui a supporté des coûts pour la produire ?

Si par croisement et analyse de données, une entreprise infère les goûts alimentaires d'un individu, ses problèmes de santé, ses opinions politiques, ses orientations sexuelles... avec une très forte probabilité → peut-elle les valoriser ?

Question de concurrence : les transactions bilatérales sont-elles le marché ?



Si les plates-formes sont capables d'adapter et d'ajuster en temps réel le prix des transactions bilatérales avec chaque consommateur (Yield management), comment vérifier que le prix est un prix de marché concurrentiel sans l'existence du « commissaire priseur walrasien » garant que le prix est bien optimal et n'est ni excessif, ni discriminatoire...

*L'Internet tend à produire le vrai marché (modèle Arrow-Debreu)
chaque transaction est unique : est-ce souhaitable ? Le contrôle est-il possible ? Est-ce socialement acceptable ?*